心理科学进展　2019, Vol. 27, No. 1, 181–189
Advances in Psychological Science　DOI: 10.3724/SP.J.1042.2019.00181

·研究方法(Research Method)·

对齐(Alignment)——一种新的多群组分析法

温聪聪 [1]　伍伟平 [1]　林光杰 [2]
([1] 厦门大学海外教育学院/国际学院, 厦门　361102) ([2] 厦门大学新闻传播学院, 厦门　361005)

摘　要　进行跨群组的因素均值比较需要检验测量工具的可比较性, 常见的做法是使用多群组 CFA 检验各群组的截距恒定性, 但截距恒定假设因其过于苛刻而往往无法满足。对齐法在这一背景下应运而生, 通过检验构置恒定模型中参数的近似恒定性而进行跨群组因素均值比较。文章介绍了多群组 CFA 和对齐法的原理, 运用某省 4 个学校的本科生职业价值观研究实例来演示如何在 Mplus 软件中使用对齐法进行研究, 并总结了对齐法的优势与局限, 回顾了对齐法近年的研究进展。

关键词　多群组分析; 多群组 CFA; 测量恒定性; 对齐法; 蒙特卡洛模拟研究
分类号　B841

1　研究背景

在传统的实证研究中, 因素分析可以将众多观测变量降维, 提取出少量的因素来解释观测变量, 让研究者对题项的分类有更清楚的认识, 也可以运用更简洁的因素得分进行进一步研究。后来, 当研究者们希望使用同一测量工具进行跨群体比较时, 发现因素分析降维所得到的维度对组成总样本的不同子样本来说含义可能有所不同, 将因素均值和因素得分直接进行比较可能存在问题。所以, 检验维度的可比较性(comparability)成为越来越受关注的问题。一些跨国研究探究了测量工具的跨群体间的可比较性, 获得了许多有价值的发现。Billiet 对欧洲各国男女性的宗教信仰进行了研究, 发现相比其他欧洲国家的女性, 土耳其女性是唯一回答参与宗教服务的频率比男性低的群组, 有些研究者可能会下“土耳其是欧洲唯一的男性比女性更信教的国家”的结论。但当及时了解到在土耳其这样的伊斯兰国家, 即使女性信教, 一般也不允许女性参与宗教服务, 作者判断这一题项对土耳其回答者的含义要不同于其他欧洲国家的回答者, 题项仅在其他欧洲国家间具备可比较性。正是发现了特殊的国情差异, 作者没有因为这一题项的异常而低估土耳其女性的信教程度(Billiet, 2013)。Piurko 等人研究了东欧和西欧国家公民的“左”和“右”的政治态度, 发现对“左”和“右”的理解可能在欧洲的自由主义国家(如瑞典)、传统主义国家(如希腊)、后共产主义国家(如捷克)间是不一样的(Piurko et al., 2011), 所以根据“左”和“右”的政治倾向把国家归类, 比较不同国家公民之间的回答可能并不合理。总之, 检验所研究维度在不同群体之间的可比较性十分重要。如果不具备可比较性, 因素分析所得出的因素得分或是因素均值就不能直接进行跨群组比较。

目前, 比较常用的一种检验维度可比较性的方法是多群组验证性因素分析(多群组 CFA, multiple-group confirmatory factor analysis)。研究者可以运用多群组 CFA 检验所得模型的测量恒定性(measurement invariance, Davidov et al., 2014; Millsap, 2011)来得出量表维度是否具备跨群体可比较性的结论。检验测量恒定性包括检验构置恒定(configural invariance 或 pattern invariance)、因素载荷恒定(metric invariance 或 weak measurement invariance)、截距恒定(scalar invariance 或 strong measurement invariance)、残差恒定(uniqueness invariance 或 strict measurement invariance)、因素方差–协方差恒定(factor variance-covariance

收稿日期: 2018-03-08
通信作者: 温聪聪, E-mail: wencong001@xmu.edu.cn

invariance)、因素均值恒定(factor mean invariance)等。构置恒定[1]表示对于不同的群组，量表所包含的因素数目和各因素设为定值和设为自由估计的因素载荷是相同的(Marsh et al., 2017; Millsap, 2011; Vandenberg & Lance, 2000; Wang & Wang, 2012; Widaman & Reise, 1997)，检验该假设需要设定模型中不同群组有相同的因素数目，相同估计模式的因素载荷；因素载荷恒定表示对于不同的群组，因素上每一单位的变化使观测变量产生的变化是相同的，检验该假设需要在设定相同因素数目和因素载荷估计模式的基础上进一步设定模型中不同群组有相等的因素载荷；截距恒定表示对于不同的群组，观测变量的初始参照值是相同的，检验该假设需要在设定相同因素结构、相同的因素载荷的基础上进一步设定模型中不同群组有相等的截距；残差恒定等以此类推。由于这些依次限定参数恒定的模型具有嵌套关系，所以检验这些测量恒定性假设需要计算所研究模型两两间的卡方变化值和自由度变化值，进而计算出显著性，得出是否拒绝原假设。通常情况下，只要不同群组的构置、因素载荷和截距都满足测量恒定，就说明不同群组对研究维度的理解相同，进而可以对不同群组的因素均值进行直接比较，探究不同群组在所得维度上的表现差异。

大多数的实证研究都希望通过多群组 CFA 探究不同群组在潜在因素上的表现差异，也就是进行因素均值比较。但多群组 CFA 最大的不足是依赖严苛的截距恒定假设才能进行群组间因素均值比较，而截距恒定假设在许多情况下无法满足。在这些情况下，多数研究者会试图借助于修正指数(modification index)释放一些参数限定来改善模型拟合，可能会得到部分截距恒定模型，继续进行均值比较。但 Asparouhov 和 Muthén 在 2014 年指出，通过修正指数进行模型修正需要人工手动选择，存在主观性。另外，通过修正指数进行模型修正可能导致得出并不正确的部分恒定模型，因为截距恒定模型本来就和真实模型相差甚远(Asparouhov & Muthén, 2014)。Marsh 等人也在 2017 年指出，虽然通过修正指数释放参数可以得到模型拟合度较好的部分截距恒定模型，但这样做并不能保证得到没有偏差的因素均值(Marsh et al., 2017)。因此，当截距恒定被拒绝时，如何继续进行组间比较一直以来都是难题。

Asparouhov 和 Muthén 在 2014 年提出了一种“对齐” (Alignment)方法，并不依赖截距恒定，只需要模型中的大多数参数满足近似恒定就可以进行因素均值比较(Asparouhov & Muthén, 2014)。对齐方法相比传统方法拥有众多优势，但可能由于对传统方法的依赖、新方法理解有一定难度、新方法只能在 Mplus 软件上实现等原因，对齐法目前只在少数国外实证研究中被使用。本文希望对新方法的原理和优势进行介绍，并运用我国某省 4 个学校的本科生职业价值观研究实例来演示如何使用对齐法进行实证研究。

2 对齐方法的原理

从最一般的线性回归模型入手，其一般形式如下式所示：

$$y = a + bx + \varepsilon \tag{1}$$

其中 y 为因变量，也可称为结果变量，对应坐标轴 y 轴上的数值。x 为自变量，也可称为预测变量，用来解释因变量，对应坐标轴 x 轴上的数值。a 为截距，是直线与 y 轴的交点的数值，此时 $x = 0$，截距值是 y 的初始参考值。b 为斜率，也就是回归系数，直接反映了自变量 x 每变化一个单位因变量 y 所改变的值。ε 为残差，代表观测值与预测值的偏离，服从正态分布。线性回归描述了自变量和因变量间最一般的线性关系。

如果把一般线性回归模型推广至因素分析，那么假设多群组 CFA 模型为下式：

$$y_{ipg} = v_{pg} + \sum_{m=1}^{m} \lambda_{pmg}\eta_{img} + \varepsilon_{ipg} \tag{2}$$

其中 y 为观测变量，属于因变量；v 为截距，与一般线性回归类似，代表观测变量初始参照值(蔡华俭，林永佳，伍秋萍，严乐，黄玄凤，2008);

[1] 国内有研究者使用“形态恒定”这一中文翻译(王孟成，2014)。但形态恒定的英文原文为 pattern invariance，与 configural invariance 有所不同。对于 configural invariance，国内许多研究者译为“结构恒定”，笔者认为这一译法并不妥当。一方面，“结构的”对应的英文为 structural，将其和 configural 等同似乎并不恰当；另一方面，如上文所述，configural invariance 代表各群组因素结构、数目相同，还代表各因素设为定值和设为自由估计的因素载荷相同，所以 configural 不仅有因素结构一致的含义，还有因素载荷配置情况一致的含义。综上所述，configural invariance 的中文译文不应当是“结构恒定”，笔者认为应当是“构置恒定”。

λ为因素载荷, 相当于一般线性回归中的斜率, 代表因素 η 每变化一个单位在相应观测变量上的变动值; η 为因素, 也就是问卷量表中各观测变量所包含的维度, 相当于一般线性回归中的自变量 x; 残差仍然为 ε。这些参数可以属于不同的群组、因素、观测变量、不同群组中的个案, 所以有了下标。其中 $p = 1, 2\cdots, p$,为观测变量序数; $m = 1, 2\cdots, m$, 为因素序数; $g = 1, 2\cdots, g$,为群组序数; $i = 1, 2\cdots, i$,为群组 g 中的独立个案序数。假设 ε_{ipg} 均值为 0, 方差为 θ_{pg}, 服从正态分布; 假设 η_{ig} 均值为 α_g,方差为 ψ_g,服从正态分布。

前文所提到的构置恒定模型就是要限制各群组 η 的数目和对应的设为定值和设为自由估计的 λ 相同, 因素载荷恒定就是要在构置恒定前提下继续限制各群组因素载荷 λ 相同, 截距恒定就是要在构置恒定和因素载荷恒定前提下继续限制各群组截距 v 相同。

如果多群组 CFA 测量恒定性检验满足了截距恒定, 就可以判定所得出的因素分析模型适用于所有群组, 维度具备可比较性, 可以直接对各群组的因素均值进行比较。但截距恒定假设所有群组的因素载荷和截距完全相等的条件过于苛刻, 在实际研究中很难得到满足, 尤其是当模型较为复杂, 群组数较多的时候。如果多群组 CFA 测量恒定性检验不满足截距恒定, 那么如上文所述, 研究者即使依靠部分截距恒定模型也很难得到正确的模型, 组间因素均值比较变得无从下手。

在多群组 CFA 构置恒定模型设定下, 因素均值和因素方差被设为定值, 无法被自由估计, 个案间的因素得分比较和群组间的因素均值比较都无法进行。只有分别限制了因素载荷恒定和截距恒定, 因素方差和因素均值才能被自由估计。与基于截距恒定模型进行组间因素均值比较的多群组 CFA 不同, 对齐法进行组间因素均值比较是基于构置恒定模型。

假设构置恒定模型为 M0, 将其截距和因素载荷分别记为 v_{pg0} 和 λ_{pg0}。M0 要求将因素转化为均值为 0, 方差为 1, 各群组需要做如下变换:

$$\eta_{g0} = (\eta_g - \alpha_g) / \sqrt{\psi_g} \tag{3}$$

同时, 为了让所有变量拥有相同的单位(metric), 方便不同观测变量进行比较, 对齐法也要求标准化所有观测变量。

观测变量的方差和均值如下式所示:

$$V(y_{pg}) = \lambda^2_{pg}\psi_g = \lambda^2_{pg,0} \tag{4}$$

$$E(y_{pg}) = v_{pg}+\lambda_{pg}\alpha_g = v_{pg,0} \tag{5}$$

因此通过公式变换可以得到:

$$\lambda_{pg,0} = \lambda_{pg}\sqrt{\psi_g} \tag{6}$$

$$v_{pg,0} = v_{pg} + \frac{\lambda_{pg,0}}{\sqrt{\psi_g}}\alpha_g \tag{7}$$

对于任意群组的 α_g 和 ψ_g, 总有对应的截距 v_{pg} 和因素载荷 λ_{pg}, 使得所得模型和 M0 具有相同的似然值, 并且由下式所得到:

$$\lambda_{pg,1} = \frac{\lambda_{pg,0}}{\sqrt{\psi_g}} \tag{8}$$

$$v_{pg,1} = v_{pg,0} - \alpha_g\frac{\lambda_{pg,0}}{\sqrt{\psi_g}} \tag{9}$$

通过赋予每个组别不同的因素均值 α_g 和因素方差 ψ_g, 利用待分析模型和最初构置恒定模型有相同模型拟合度的关系计算出对应的因素载荷 λ 和截距 v, 如此不断重复、迭代, 直到以下总衰减函数(total loss function) F 被最小化, 模型的参数不恒定性也被最小化, 此时就找到了最优测量恒定模型。

$$F = \sum_p \sum_{g1<g2} W_{g1,g2} f\left(\lambda_{pg1,1} - \lambda_{pg2,1}\right) + \sum_p \sum_{g1<g2} W_{g1,g2} f\left(v_{pg1,1} - v_{pg2,1}\right) \tag{10}$$

对于任意一对群组 $g1$、$g2$, 截距和因素载荷的差异通过成分衰减函数(component loss function) f 的缩放被纳入进总衰减函数 F 进行计算。在 Mplus 中使用对齐法分析时所使用的成分衰减函数 f 如下式所示:

$$f(x) = \sqrt{\sqrt{x^2 + \varepsilon}} \tag{11}$$

此处 ε 为一个很小的正值, 保证了 f 具有一阶连续导数。利用 f 的一阶连续导数, 总衰减函数 F 的优化过程(optimization process)变得更容易, 方便了总衰减函数 F 的计算。成分衰减函数 f 约等于 $\sqrt{|x|}$。总丢失函数 F 取得最小值解将是当模型中有少数较大的不恒定参数和大部分近似恒定的参数的时候。此时, 大部分参数近似恒定, 不同群组相应参数的差异 x 接近为 0, $f(x)=\sqrt{|x|}$, 取得最小值; 对于少数较大的不恒定参数, 需要考虑两种情况: 当 $x<1$ 时, $f(x)>x$, 参数不恒定性被放大, 较大的不恒定参数比较小的不恒定参数放大模型

不恒定性的程度更小；当 $x>1$ 时，$f(x)<x$，参数不恒定性得到衰减，较大的不恒定参数比较小的不恒定参数使模型不恒定性衰减的程度更大。所以，根据 f 的特点，x 接近零本身取得最小值。如果必须有少量的 x 不接近零，那么 x 应当尽量大。因为 x 越大，函数衰减的越多，模型的不恒定性衰减的越多。这和 EFA 中的旋转函数保留较大和较小载荷，排除中等载荷的特点类似。

在公式(10)中，总衰减函数 F 中的 $W_{g1,g2}$ 是用来反映群组大小和特定群组参数估计的确定性。较大的群组比较小的群组对总丢失函数的贡献更大。具体计算如下式所示：

$$W_{g1,g2}=\sqrt{N_{g1}N_{g2}} \tag{12}$$

其中 N_{g1} 和 N_{g2} 为任意一对群组 $g1$、$g2$ 的样本量。

当总衰减函数 F 被最小化后，由于有一个群组被设为了方差为 1 的参照组(在 Mplus 中为第一组)，所以可以得到除了参照群组外的所有因素均值和因素方差和参照群组中的因素均值，共计 $2g-1$ 个参数。为了计算参照群组中的因素方差相对值，需要利用以下参数限定，令 $\psi_1=(\psi_2\times\cdots\times\psi_g)^{-1}$：

$$\psi_1\times\psi_2\times\cdots\times\psi_g=1 \tag{13}$$

在这里，对齐法又可以分为固定优化和自由优化两种算法，其区别是固定优化假设参照群组的均值 α_1 为 0，自由优化则将 α_1 自由估计。Asparouhov 和 Muthén 所做的模拟研究显示，当群组数目为 2 或是模型中没有不恒定参数时，固定优化算法要优于自由优化算法；在群组数目大于 2 并且模型不恒定时，自由优化算法要优于固定优化算法(Asparouhov & Muthén, 2014)。

3　案例应用

对齐法作为一种新方法，在实证研究中的使用并不多。笔者根据研究经验，针对在 Mplus 软件[2]中使用对齐法分析数据的步骤进行了一个总结。总得来说，使用对齐法可以遵循以下四个步骤：

1)首先要运行 EFA，检验分析出的量表结构是否和设计时的结构相同，如果不同，则证明量表设计有问题。

2)其次进行多群组 CFA 分析，检验各群组间的测量恒定性。如果模型满足截距恒定，则该模型是较为简洁的模型，已经可以进行群组间因素均值比较，没有必要再使用对齐法；如果模型拒绝截距恒定，则可以考虑使用对齐法。

3)如果需要使用对齐法，则根据所研究数据的实际情况选择固定优化算法或是自由优化算法进行分析。

4)根据研究需要进行蒙特卡洛模拟研究。此时进行模拟研究的目的可以是为了选择适当的优化算法，可以是检验对齐法估计各个参数的有效性，也可以是检验参数近似恒定性等。

同时，笔者也提出一些使用对齐法分析的注意事项：当群组数较多，模型较为复杂时，准确的对齐法参数估计要求有较大的样本量(Asparouhov & Muthén, 2014)。当模型中不恒定的参数比率达到 25%时，需要进一步使用模拟研究检验对齐法的有效性(Asparouhov & Muthén, 2014)。如果需要进行模拟研究，则需要注意在 Mplus 中使用 svalues 指令保存对齐法计算出的参数估计结果，作为模拟研究的真实值。如果模型中的大多数参数不恒定性非常强，无法满足近似恒定，对齐法也不会有较好的分析效果。

接下来，将运用某省 4 个学校的本科生职业价值观研究实例[3]来演示当多群组 CFA 模型不满足截距恒定假设时，应当如何使用对齐法检验模型参数的近似恒定性，进而进行因素均值比较。参与本次调研的 4 个学校的本科生总样本共计 4692 名学生，其中 A 学校 2729 名，B 学校 615 名，C 学校 455 名，D 学校 893 名。4 个学校各自的样本量较多，可以满足对齐法分析的大样本量要求。研究共分为 4 个步骤，供读者参考。

3.1　EFA 分析

本研究使用的职业价值观量表是印第安纳大学问卷(Assessing Multinational Interest in STEM)的简化修改版(Maltese et al., 2014)，如表 1 所示。量表预设为个人兴趣、社会责任和经济动机三个维度，使用 4 点量表(1.不重要、2.不太重要、3.比较重要、4.重要)的形式进行测量，设问为“考虑

[2] Mplus 7.1 以上的软件版本都提供了使用对齐法分析数据的功能。

[3] 本研究实例的数据和相关 Mplus 语法文件已经上传至 OSF 网站供读者参阅。项目链接为：https://osf.io/k34sg/

到您自己的未来发展, 您认为下列因素的重要程度如何”。

表 1　问卷中职业价值观量表的维度设计

维度设计	题号	题项内容
社会责任	Q1	有助于可持续发展和环境保护
	Q5	帮助他人
	Q9	做对社会发展有重要意义的事情
个人兴趣	Q2	自我发展
	Q3	做我感兴趣的事情
	Q4	获得一份有保障的工作
	Q6	发挥我的天赋和能力
经济动机	Q7	获得高薪的机会
	Q8	尽快开始赚钱

使用 Mplus 软件运行 EFA 后, 所得的表 2 所示的旋转后因素载荷矩阵结构和问卷的维度设计完全一致, 量表设计得到了数据样本的支撑。

表 2　旋转后的因素载荷(N = 4692)

题项	F1	F2	F3
1. 有助于可持续发展和环境保护		**0.767**	
2. 自我发展	**0.742**		
3. 做我感兴趣的事情	**0.833**		
4. 获得一份有保障的工作	**0.625**		
5. 帮助他人		**0.828**	
6. 发挥我的天赋和能力	**0.689**		
7. 获得高薪的机会			**0.671**
8. 尽快开始赚钱			**0.705**
9. 做对社会发展有重要意义的事情		**0.717**	

注: *黑体数值在 5%的置信区间上显著

3.2　多群组 CFA 分析

分别限制模型构置恒定、因素载荷恒定、截距恒定, 使用 Mplus 软件运行多群组 CFA 测量恒定性检验, 得到表 3 所示的模型拟合结果。构置恒定模型的卡方值最低, 表明拟合度最好。经过嵌套模型计算后, 卡方变化值与自由度变化值对应的 p 值在 5%的置信区间上拒绝截距恒定和因素载荷恒定模型,构置恒定模型是最优模型。由于截距恒定模型被拒绝, 因此可以继续使用对齐法检验近似测量恒定性, 进而尝试进行因素均值比较。

表 3　使用多群组 CFA 进行测量恒定性检验得到的模型拟合结果(N = 4692)

模型	参数数目	卡方值	自由度	p 值
构置恒定	120	1028.009	96	<0.001
因素载荷恒定	102	1058.646	114	<0.001
截距恒定	84	1120.288	132	<0.001
载荷恒定对构置恒定	–	30.638	18	0.0317
截距恒定对构置恒定	–	92.279	36	<0.001
截距恒定对载荷恒定	–	61.614	18	<0.001

3.3　对齐法分析

分别使用固定优化算法和自由优化算法在 Mplus 软件中执行对齐法分析, 得到以下参数近似恒定结果和因素均值估计结果。(后文通过模拟研究证明固定优化算法得出的结果较为可靠, 此处仅展示部分固定优化算法得到的参数估计结果)从表 4 所示的参数近似恒定结果来看, 所有截距和因素载荷都近似恒定, 对齐法得到了理想的结果, 群组间因素均值比较切实可行。而在上一步使用多群组 CFA 时, 因素载荷恒定模型和截距恒定模型都被拒绝, 因素均值比较无法进行。从这一方面来看, 对齐法克服了群组间因素载荷和截距完全恒定的苛刻假设, 通过检验因素载荷和截距参数的近似恒定性使不同群组因素均值间的直接比较成为可能, 这是其相较多群组 CFA 的一大优势。从表 5 所示的个人兴趣因素均值比较来看, 群组 4 显著高于群组 1,2,3; 从表 6 所示的经济因素均值比较来看, 群组 2 和群组 3 显著高于群组

表 4　参数近似恒定结果(N = 4692)

参数类型	近似恒定组别	参数类型	近似恒定组别
截距		因素载荷	
Y1	1,2,3,4	Y1	1,2,3,4
Y2	1,2,3,4	Y2	1,2,3,4
Y3	1,2,3,4	Y3	1,2,3,4
Y4	1,2,3,4	Y4	1,2,3,4
Y5	1,2,3,4	Y5	1,2,3,4
Y6	1,2,3,4	Y6	1,2,3,4
Y7	1,2,3,4	Y7	1,2,3,4
Y8	1,2,3,4	Y8	1,2,3,4
Y9	1,2,3,4	Y9	1,2,3,4

1,4。考虑到 4 所院校层次的排序为 4>1>2≈3，对齐法的研究结果与之前全国样本“985 院校学生相比 211 院校和普通本科院校学生更注重个人兴趣；普通本科院校学生相比 985 院校学生和 211 院校学生更注重经济因素”的研究结论相一致(温聪聪, 伍伟平, 蒋玉塔, 2016)。

表 5 个人兴趣因素的均值比较结果(N = 4692)

排名	群组号	因素均值	因素均值显著小于该组
1	4	0.162	2,3,1
2	2	0.038	
3	3	0.028	
4	1	0.000	

表 6 经济因素的均值比较结果(N = 4692)

排名	群组号	因素均值	因素均值显著小于该组
1	2	0.163	1,4
2	3	0.127	1,4
3	1	0.000	
4	4	–0.014	

3.4 蒙特卡洛模拟研究

在第三步对齐法分析中并不清楚应该使用固定优化算法还是自由优化算法比较好，所以在这一步中，应当分别利用两种优化算法得到的参数估计结果进行蒙特卡洛模拟研究，探究使用特定算法的对齐法能否较好分析参数间的恒定与不恒定性，能否无偏估计各个参数。本研究中检验使用不同算法的对齐法是否无偏估计各个参数的标准有两个，一是相应算法的因素均值真实值和估计值的比较，体现因素均值出现了多少数值的偏差。二是相应算法因素均值的均方误差，通过将该因素均值的方差和因素均值的期望偏差的平方加和得到，体现参数估计的效果。均方误差越小越好，且当等于所估计参数的方差的时候说明是无偏估计。模拟研究的条件基本模仿真实情况，选择 4 个群组，与真实数据相同的群组样本量(2729、615、455、893)，生成数据组数为 500。使用第三步固定优化算法得到的参数估计结果进行固定优化对齐分析，使用自由优化算法得到的参数估计结果进行自由优化对齐分析，得到以下因素均值估计结果和因素均方误差计算结果。从表 7 所示的因素均值估计结果来看，使用固定优化算法的对齐法能较好估计各群组的因素均值；但使用自由优化算法的对齐法因素均值估计出现了较大偏差，尤其是在经济因素，偏差超过了 1.3。从表 8 所示的因素均方误差计算结果来看，两种算法在因素方差上的均方误差完全相同，但自由优化算法在各因素均值上的均方误差都要大于固定优化算法。尤其是在经济因素，自由优化

表 7 两种算法的因素均值估计结果(N=4692)

均值类别	固定优化算法真实值	估计值	自由优化算法真实值	估计值
群组 1				
个人兴趣因素	0.000	0.000	0.314	0.221
社会责任因素	0.000	0.000	–0.081	–0.072
经济因素	0.000	0.000	1.848	0.520
群组 2				
个人兴趣因素	0.038	0.040	0.351	0.261
社会责任因素	0.075	0.079	–0.004	0.008
经济因素	0.163	0.164	2.009	0.682
群组 3				
个人兴趣因素	0.028	0.033	0.339	0.252
社会责任因素	0.065	0.070	–0.018	–0.002
经济因素	0.127	0.129	1.975	0.648
群组 4				
个人兴趣因素	0.162	0.163	0.472	0.381
社会责任因素	–0.005	–0.006	–0.086	–0.077
经济因素	–0.014	–0.021	1.810	0.484

表 8 两种算法的因素均方误差计算结果(N = 4692)

因素	固定优化算法均方误差		自由优化算法均方误差	
	均值	标准差	均值	标准差
个人兴趣因素				
因素均值	0.036	0.016	0.224	0.179
因素方差	0.046	0.020	0.046	0.020
社会责任因素				
因素均值	0.038	0.017	0.184	0.141
因素方差	0.050	0.022	0.050	0.022
经济因素				
因素均值	0.041	0.018	1.527	1.139
因素方差	0.060	0.025	0.060	0.025

算法估计因素均值的均方误差均值达到 1.527, 而固定优化算法的这一数值仅为 0.041。综上所述, 在本部分模拟研究所设定的研究条件下, 使用固定优化算法的对齐法可以几乎无偏估计因素均值, 均方误差较小, 效果明显优于使用自由优化算法的对齐法。回想第三步中所有因素载荷和截距都近似恒定, 这一结论和 Asparouhov 和 Muthén 所做模拟研究中“当模型中没有不恒定参数时, 固定优化算法要优于自由优化算法”的结论相一致(Asparouhov & Muthén, 2014)。所以, 第三步仅列出了使用固定优化算法的对齐法计算出的参数估计结果。

4　方法总结与展望

通过前文所述, 对齐法可以被认为是传统多群组 CFA 的有益补充, 其最大贡献是可以克服苛刻的截距恒定假设, 通过检验参数的近似恒定性而达到群组间的因素均值比较的目的。相比传统的多群组 CFA, 对齐法可以有效拟合参数近似恒定的模型, 模型拟合度和拟合度较好的构置恒定模型完全相同, 还可以处理数目繁多的群组、较为复杂的量表结构, 可以为模型中因素载荷和截距的近似恒定性提供详细的诊断。通常情况下, 如果研究数据的样本量很大, 群组较多, 模型较复杂, 那么使用多群组 CFA 很难满足截距恒定假设。此时, 使用对齐法检验模型参数近似恒定性而进行因素均值比较可能是较好的选择。如果研究者希望了解模型中因素载荷和截距等参数的恒定性, 从而对模型进行诊断或是希望改进量表设计, 使用对齐法也是更好的选择, 因为传统的使用修正指数释放参数的方法受研究者主观判断的影响较大, 修正指数的多重共线性也会误导研究者的判断。

同时, 对齐法作为一种新方法, 也存在着一些局限性。首先, 对齐法是基于构置恒定模型, 潜在的前提是构置恒定模型拟合较好, 且要优于因素载荷恒定模型或是截距恒定模型。但有的时候会出现构置恒定模型本身拟合就很不好或是因素载荷恒定模型拟合更好, 此时使用对齐法就会有一定的问题。其次, 对于较多的群组和较为复杂的模型, 对齐法要求待分析的各个群组的样本量足够大, 而这一条件不一定总能满足。再次, 对齐法的提出者仅仅指出当模型中的不恒定参数比率小于等于 25%时对齐法参数估计能取得较好效果(Asparouhov & Muthén, 2014), 但 25%的依据并不充分, 今后的研究可以探究这一比率是否更小或更大。最后, 对齐法仍然是基于传统 CFA, 只允许观测变量在一个因素上有因素载荷, 不允许因素拥有跨因素载荷(cross-loadings), 所以研究的 CFA 模型可能本身就不是正确的模型。

对齐法自 2014 年提出以来也在不断更新, 方法被应用于更多领域, 也有许多新进展。对齐法提出同年, Muthén 和 Asparouhov 将对齐法应用于项目反应理论(item response theory)研究(Muthén & Asparouhov, 2014)。他们介绍了在项目反应理论情境下对齐法的计算方法, 运用蒙特卡洛模拟研究检验了对齐法的适用性。2017 年, Muthén 和 Asparouhov 比较了对齐法和随机效应法(random effects method)两种研究多群组测量恒定性的方法, 详细叙述了在不同研究条件下选取对齐法或是随机效应法分析数据的准则, 比较了两种方法各自的优势与缺陷。随机效应法的引入进一步填补了测量恒定性研究领域的空白, 是对齐法的有益补充。2017 年, Marsh 等人基于标准对齐法提出了 CFA 对齐法(alignment within CFA), 将更倾向于探索性的对齐法变为了验证性的对齐法(Marsh et al., 2017)。相比标准对齐法, CFA 对齐法可以更广泛地应用于结构方程模型, 引入协变量进行多指标多因素(MIMIC)模型分析, 并且在 Mplus 软件中还可以得到标准对齐法无法提供的修正指数等指标。

随着研究者越来越深入了解对齐法的原理, 有越来越多的研究开始使用对齐法进行多群组分析, 研究群组间的可比较性。Munck 等人使用对齐法研究了 1999 年教育成就评价国际委员会收集的欧洲 22 国公民教育研究数据和 2009 年该机构收集的欧洲 24 国国际公民与公民身份教育研究数据(Munck & Torney, 2017)。研究者以国别和性别两个变量区分不同群组, 运用对齐法研究了 92 个研究组别的测量恒定性, 成功比较了样本中国别与性别的差异, 解决了没有满足截距恒定无法进行群组间因素均值比较的问题。Lomazzi 使用对齐法研究了世界民众价值观调研收集的关于男女性承担角色的态度的全球 59 国数据样本(Lomazzi, 2018), 比较了部分截距恒定模型和对齐法各自的特点和优势。

最后，基于对齐法的局限性，可能仍然有许多等待研究者挖掘的领域。例如，不同情境下对齐法需要的群组样本量大小，对齐法取得较好估计效果所能接受的不恒定参数率，对齐法与探索性结构方程模型(ESEM)结合的可能性，运用对齐法研究跨时序模型中参数近似恒定性的可能性，对齐法和多水平随机效应法的适用性等等。

参考文献

蔡华俭，林永佳，伍秋萍，严乐，黄玄凤．(2008)．网络测验和纸笔测验的测量不变性研究——以生活满意度量表为例．*心理学报, 40*(2), 228–239.

王孟成．(2014)．*潜变量建模与 Mplus 应用•基础篇*．重庆大学出版社．

温聪聪，伍伟平，蒋玉塔．(2016)．厦门大学理工科本科生职业发展研究．*职大学报,* (3), 94–99.

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling A Multidisciplinary Journal*, *21*(4), 495–508.

Billiet, J. (2013). Quantitative methods with survey data in comparative research. *Urology*, *76*(76), S2–S3.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(40), 55–75.

Lomazzi, V. (2018). Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology (MDA)*, *12*(1), 77–103.

Maltese, A. V., Ross, H., Wang, L., & Wang, Y. (2014). Assessing multinational interest in STEM: Implementing a comparative survey research study in China. *International Journal of Chinese Education*, *3*(1), 109–131.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., & Muthén, B., et al. (2017). What to do when scalar invariance fails: The extended alignment method for multiple-group factor analysis comparison of latent means across many groups. *Psychological Methods*. Advance online publication.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY, US: Routledge/Taylor & Francis Group.

Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology*, *5*(5), 978.

Muthén, B., & Asparouhov, T. (2017). Recent methods for the study of measurement invariance with many groups: alignment and random effects. *Sociological Methods & Research*, DOI: 10.1177/0049124117701488.

Munck, I., Barber, C., & Torney-Purta, J. (2017). Measurement invariance in comparing attitudes toward immigrants among youth across Europe in 1999 and 2009: The alignment method applied to IEA CIVED and ICCS. *Sociological Methods & Research*, DOI: 10.1177/0049124117729691.

Piurko, Y., Schwartz, S. H., & Davidov, E. (2011). Basic personal values and the meaning of left-right political orientations in 20 countries. *Political Psychology*, *32*(4), 537–561.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.

Wang, J. C., & Wang, X. Q. (2012). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC, US: American Psychological Association.

Alignment: A new method for multiple-group analysis

WEN Congcong[1]; WU Weiping[1]; LIN Guangjie[2]

([1] *Overseas Education College/International College, Xiamen University, Xiamen* 361102, *China*)

([2] *School of Journalism and Communication, Xiamen University, Xiamen* 361005, *China*)

Abstract: Comparing the latent factor means across groups need to test the comparability of the instrument. Researchers usually test the scalar invariance to achieve this goal but often get unsatisfactory results. The alignment is proposed under this background. It allows the comparison of the latent factor means by testing the approximate invariance of the configural model parameters and has become a significant complement to multiple-group CFA. This article gives a detailed description of the multiple-group CFA and alignment, summarizes the research procedures and points to which researchers need to pay attention when applying alignment, uses an undergraduates' work value data to illustrate how to use alignment to do a research with Mplus. In the end, this article summarizes the advantages and limitations of alignment and reviews its research developments and empirical applications.

Key words: multiple-group analysis; multiple-group CFA; measurement invariance; alignment; Monte Carlo simulation study